DOCUMENT RESUME

ED 396 315 CS 215 359

AUTHOR	Calfee, Robert
TITLE	Implications of Cognitive Psychology for Authentic Assessment and Instruction. Technical Report No. 69.
INSTITUTION	National Center for the Study of Writing and Literacy, Berkeley, CA.; National Center for the Study of Writing and Literacy, Pittsburgh, PA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	May 96
CONTRACT	R117G10036
NOTE	25p.
PUB TYPE	Viewpoints (Opinion/Position Papers, Essays, etc.) (120)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Accountability; *Behaviorism; *Cognitive Psychology; Educational Trends; Elementary Secondary Education; Higher Education; *Learning Theories; Testing
IDENTIFIERS	*Authentic Assessment; *Educational Issues

ABSTRACT

Designed to develop a conceptual perspective, this paper focuses on three tensions: between cognitive and behavioral views of learning and thinking; between factory-model and information-age models of schooling; and between externally-mandated testing and internally-guided assessment. The paper first provides a brief sketch of developments in the psychology of learning and thinking over the past half century. The paper then presents a few thoughts about the fork in the road that now confronts United States educators, the path of least resistance continuing a tradition of "managed" schooling, and the more challenging path calling for a radical transformation in the teaching profession. The third section of the paper focuses on testing and assessment, probably the point of greatest tension. The paper concludes with a description of an assessment model that relies on teacher judgments for both internal and external accountability. Contains 5 figures and 31 references. (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 396 315

National Center for the Study of Writing

Technical Report No. 69

Implications of Cognitive Psychology for Authentic Assessment and Instruction

Robert Calfee

May, 1994

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

University of California at Berkeley

Carnegie Mellon University

BEST COPY AVAILABLE

CS215359

Technical Report No. 69

Implications of Cognitive Psychology for Authentic Assessment and Instruction

Robert Calfee

May, 1994

NATIONAL CENTER FOR THE STUDY OF WRITING

University of California
Berkeley CA 94720
(510) 643-7022

Carnegie Mellon University
Pittsburgh PA 15213
(412) 268-6444

The publication of this report was supported under the Educational Research and Development Center Program (R117G10036 for the National Center for the Study of Writing) as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

This publication was produced on an Apple Macintosh IIsi computer with portrait display monitor and an Apple LaserWriter IIntx printer donated to the National Center for the Study of Writing by Apple Computer, Inc.

Implications of Cognitive Psychology for Authentic Assessment and Instruction

Robert Calfee
Stanford University

A Tale of Tensions

This contribution reflects personal experiences and limitations. On the one hand, the topic is children and youth, and yet I am clearly a "grown up." The audience is international, and yet I bring an admittedly U. S. perspective to the topic. The audience comprises "policy makers, scholars, and practitioners," and it is unclear where a cognitive psychologist fits into this group, if anywhere. Finally, the International Commission is concerned with testing, yet my interests lean toward learning and teaching—the latter mainly as it has to do with the former.

The title, slightly altered from the original assignment (will points will be taken off?), covers a lot of territory. For guidance through this treacherous terrain, I will rely on three overarching tensions:

- The tension between cognitive and behavioral views of learning and thinking;
- The tension between *factory-model* and *information-age* models of schooling;
- The tension between *externally-mandated* testing and *internally-guided* assessment.

My text has a historical-narrative flow, partly because a story is easier to remember than an exposition, partly because that is the way my thinking has evolved. This summary highlights trends, and so neglects significant details.

In the late 1950's when I began graduate work in psychology, behaviorism was at its peak, affecting theory and research in the behavioral sciences, and influencing practice in education, personnel selection, and other fields of human endeavor. By the time I completed my degree in 1963, the cognitive revolution was in full swing. Experimental psychology had refocused on mental processing, on memory, thinking, problem-solving, psycholinguistics, on the mysteries of the mind.

At the same time, as Cronbach (1975) noted, the breach between the psychology of learning and the psychology of testing had reached a point of no return; the cognitive revolution affected mainly psychologists in the first camp. Standardized testing procedures, arguably the most significant

accomplishment of educational psychology, had become the norm for assessing student achievement, both for administrative accountability (to be sure, parents were still interested in what teachers had to say about their students), but increasingly to guide micro-level instructional decisions (the apogee reached in computer-assisted instruction). These tests embodied behavioral principles; items were designed to assess mastery of specific performance objectives.

Neither cognitive psychology nor educational psychology was grounded in classroom practice. The former experimented with college students in laboratory settings; the latter studied computer printouts and Pearson product-moment correlations. The individual learner was "error variance" for the experimentalist, and a "normal curve equivalent" for the psychometrician. Teachers and classrooms were not in the picture, except for studies of low-inference classroom behavior as correlates of standardized achievement.

Mine is an American story. Others from different contexts would tell different tales. Nonetheless, I believe that this story and these themes spotlight issues of broad, international, and multi-cultural significance. A society's purpose in educating the young is important in deciding policy at all levels. In the United States, for example, we are committed to both quality and equality, to opening the highest levels of achievement to all children without regard to background. We still have a long way to go to achieve this aspiration. Indeed, not everyone shares this commitment, and some believe it impossible to attain.

The story includes several characters. Behaviorists appear in several guises, along with Cognitivists, both plain and "meta" flavored. Psychometricians stand as stern judges, possessors of the mystical wisdom of KR-20s, able to correct for attenuation (and other sins), and capable of Rasch analyses; they also establish validity, the possession of value. Policy makers complete the triumvirate; at the upper reaches, they are legislators and administrators, and at the bottom they become bureaucrats. The cast also includes the ephemeral troops in the trenches: schoolteachers, head masters, and students.

The episodes stretch from a time when, in the U. S., at least some teachers taught at least some students to think effectively, through decades in which students mastered behavioral objectives by repeated practice and testing, to the present, where the stage seems right for a de Maupassant ending. I admit in advance that I am not sure about the ending. In one scenario teachers regain control of curriculum and instruction, informed by a half-century of research on the psychology of thinking, their classroom judgments valued on a par with psychometric instruments. Other scenarios have less appeal.

And so, on with the story. First a brief sketch of developments in the psychology of learning and thinking over the past half century. Next a few thoughts about the forks in the road that now confront U. S. educators, the path of least resistance continuing a tradition of "managed" schooling, the more challenging path calling for a radical transformation in the teaching

profession. The third section focuses on testing and assessment, probably the point of greatest tension at present. Finally I describe an assessment model that relies on teacher judgments for both internal and external accountability. This essay is designed primarily to develop a conceptual perspective, and I will not attempt an extensive literature review. In addition to the citations supporting specific points, readers can call upon various handbooks (the three editions of the *Handbook of Research on Teaching*, Travers, 1973, Gage, 1963, and Wittrock, 1986, parallel the history presented in the next section; the *Handbook of Educational Psychology*, Berliner & Calfee, in press, provides detail on many of these points; no "Handbook of Educational Assessment" exists at present, but probably should).

From Behaviorism to Cognition in Three Easy Steps

The three panels of Figure 1 guide this first episode. The brevity of behaviorism in the top panel arises partly from an assumption that readers are familiar with tenets and research in this field, and partly from the conceptual simplicity of the area. As played out in learning theory and applications to schooling, the strategy is the decomposition of a complex task into specifiable stimulus objectives sequenced for practice, testing, and reinforcement. The basic principle works well for the acquisition of skilled tasks where transfer and reflection are not critical outcomes. To be sure, during the behavioral era, some remarkably "cognitive" work appeared, including the arena of school learning (e.g., Brownell, 1948).

By the 1970's, cognitive psychology emerged as the dominant paradigm among U. S. psychologists. As shown in the middle panel of Figure 1, stimulus and response remained in the picture, but the "organism" had become an information-processor. The computer metaphor (created in the image of man?) legitimized investigations of human thought and language, reaching a peak in the 1980's; Greeno (1980) relates the history of this paradigm shift, and I explored the implications for educational practice at about the same time (1981). The emphasis in the early stages of cognitive psychology was the study of *short-term memory*, a simple construct at first glance, but one that led to the discovery of a complex array of interrelated memories handling attention, language, analysis and interpretation. Long-term memory took shape first as a large warehouse for storing experiences, but this image quickly began to change:

> The human memory seems to be not at all like a storeroom, a library, or a computer core memory, but rather presents a picture of a complex, dynamic system... In fact, human memory does not, in a literal sense, store anything; it simply changes as a function of experience. (Estes, 1980, p. 68)

In the 1990's, cognitive psychology takes shape as the image in the lower panel of Figure 1. This picture is complex partly to make several points, but it

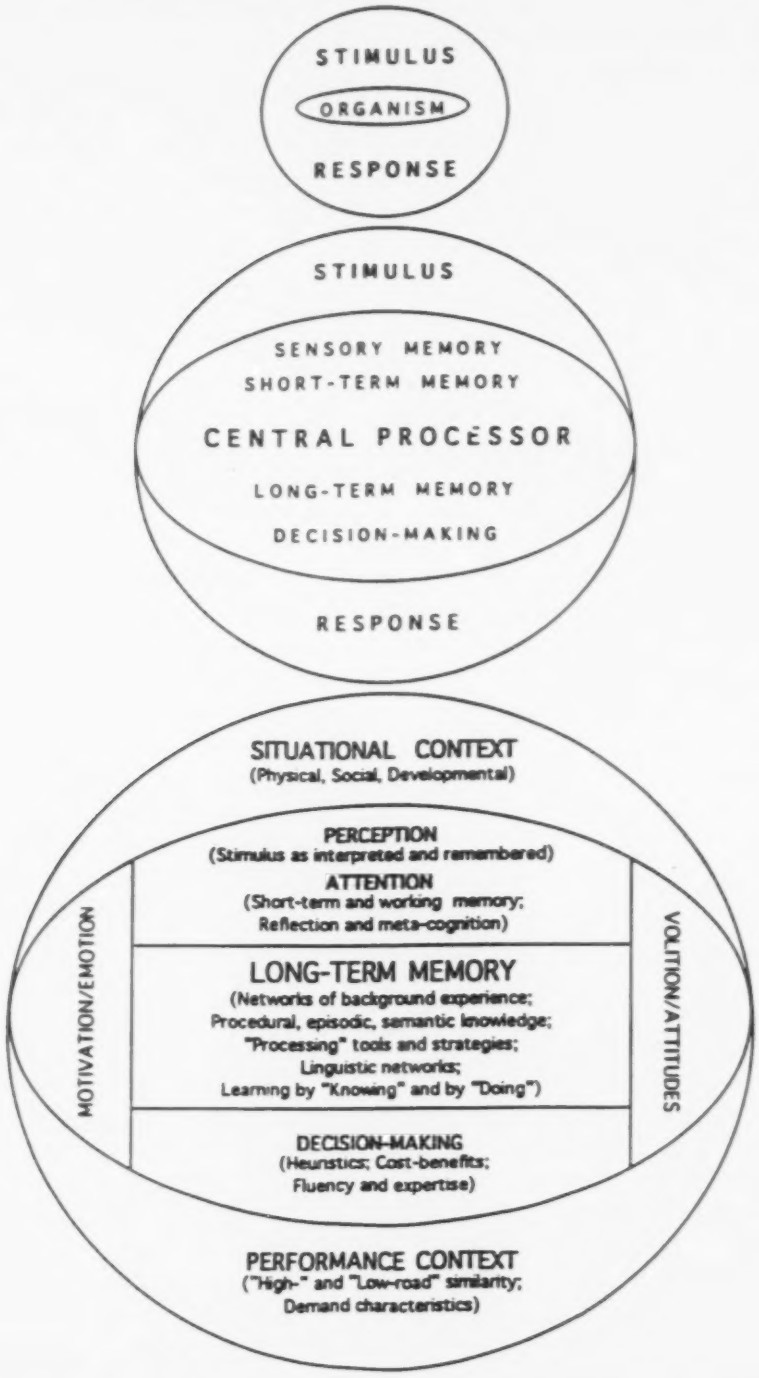

Figure 1. Developments in conceptions of learning and thinking from Behaviorism through Information-Processing to Situated Cognition, 1930-1990.

also captures the increased richness of the field. These developments may seem straightforward to Europeans more comfortable with interdisciplinary thinking, but it is a virtual revolution for U. S. Cognitivists to interact with anthropologists, ethnographers, and other "fuzzy thinkers."

Let me highlight selected features in the diagram. Short-term memory remains under the headings of perception and attention, but *long-term memory* is now center stage. The model emphasizes *categories of knowledge* (narrative images, "how to do it" routines, abstract categorical "stuff" that results from schooling), the *interplay of language and thought, strategic and dynamic "knowing" and "doing"* (computers cannot really "reflect," computer thought is not linked to action as in human beings, and so the original metaphor led scholars to overlook the constructivist aspects of cognition), and a new appreciation of the potential implications of *meta-cognition* (the term appeared in the 1960's, but began to flower in the 1980's). Long-term memory remains something of a puzzle. Some theorists characterize it as an enormous assemblage of associational pairings that communicate in parallel, while other scholars emphasize structural networks. The first view stresses the underlying randomness of memory connections, while the second focuses on the organizational features of the human mind. The warehouse metaphor offers promise in understanding this dichotomy, it seems to me. On the one hand, the first-time visitor to a warehouse (or a flea market) is thoroughly confused; the experienced aficionado sees the chaos quite differently (Chi, Glaser, & Farr, 1988). The warehouse metaphor suggests that both perspectives may have validity.

Stimulus and response have new meanings and new roles in the understanding of human thought. No longer are these elements defined to be operationally convenient. Instead, following the lead of anthropologists and social scientists, they have become challenges for analysis. Stimulus as *situated context* (Brown, Collins, & Duguid, 1989) incorporates the entire array of circumstances that affect the individual; the individual remains the focus for the cognitive psychologist, but with a new appreciation that the individual cannot be genuinely understood outside of the context. Likewise, on the response side of the equation, response as *performance* (Snow, in press) has become the code word for a broader examination of the individual's total reaction to a situation. Specifiable behaviors are still part of the equation, but the cognitivist is also likely to record qualitative facets of performance, and to ask questions like "What are you doing and why are you doing it?" Transfer has reappeared in new garb, transcending the earlier debates about specific versus general application of previous learning in a new situation; the conditions of original learning and the context of a novel situation are critical in determining whether transfer takes the high road or the low road, but both are possible (Novick, 1990; Perkins & Salomon, 1988).

Finally, affective and attitudinal elements are now in the picture. They were there before, of course, but as somebody else's problem. Now one can find serious discussions of how "skill and will" jointly influence thought and behavior, and terms like "will power" have currency among cognitive

psychologists. Snow and Jackson (1992), for instance, revived the concept of *conation* as a form of meta-motivation, the sense that individuals can reflect on their needs and goals; they describe "a wish as essentially a value attached to a goal." The analyses serve conceptually to build bridges between cognition and motivation; they help practically in suggesting how teachers can deal with the "B" word—boring.

Paradigms of Schooling

The second episode is organized around the following two questions:

- What might these shifts in our knowledge mean for school learning and achievement testing in the United States?
- What has been the impact of these ideas on actual classroom practice?

To address these questions, I will rely on Figure 2. In the United States, the marriage in the early 1900's between Educational Administration and Behavioral Psychology led to the emergence of the factory model. Unlike the English tradition of the "head teacher," U. S. principals began to "manage instruction." Their job was to keep the assembly line humming, make sure that students move through the curriculum objectives, monitor outcomes, and keep the teacher-workers on schedule.

The factory-school model is coherent and consistent. The behavioral model serves to define the *curriculum*; experts divide a complex task (e.g., reading) into a large collection of specific behaviors, which are packaged as textbooks, tests, and teachers' manuals. Students acquire each behavioral objective by practice with feedback. Student differences are handled by adjusting the students; faster students move more quickly and slower students are delayed, but the path is the same for all. *Instruction* is prescripted in the teacher's manual to follow a sequence of presentation, recitation, evaluation, and reinforcement. The *teacher's role* is to manage these activities as efficiently as possible.

The increasing frustration of U. S. policy makers with stagnant school achievement has generated frantic efforts to improve the current model. "Higher standards, longer days, greater productivity" are hallmarks of this effort, but at root the instructional assumptions undergirding the "New American Schools" are fundamentally unchanged from the factory model (Mecklenburger, 1992). The most convenient policy lever for increasing productivity in this model is the standardized multiple-choice test: cheap, mass producible, easily aggregated and quantified, and amenable to central control.

Information-age education differs in fundamental ways from the factory model. Precursors appear in Dewey's progressive education, inter alia, but the practice has seldom flourished in American schools. Two recent developments have brought this model back into the spotlight. The first is

INDUSTRIAL SOCIETY/ FACTORY SCHOOLS	INFORMATION/ INQUIRING SCHOOLS
Curriculum	
Basic skills, functional literacy	Transferrable skills, critical literacy
Separate subjects: reading, writing, arithmetic, science, history	Integrated subjects: communication and problem-solving applied to arts and sciences
Pre-specified body of knowledge, information to be memorized, emphasis on content	Emerging knowledge, strategic approach to information analysis, emphasis on process
Print-based, standard textbooks and work-sheets, "school" materials	A variety of technologies, including texts, electronic libraries, multi-media sources, "real" information from outside school
Instruction	
Teacher directed, student recitations	Teacher as facilitator of student learning and production
Individual work based on uniform processes and outcomes	Cooperative learning, group framing and solving authentic problems
Student is recipient of information; teacher is the source	Student as constructor of meaning; teacher as guide to resources
Uniform pacing for entire class or ability groups; micro-management of objectives	Pacing accommodated to student needs and interests; framed by long-term goals
Assessment	
Standardized tests; recognition and "fill in blank"	Performance-based assessments, emphasis on production of authentic projects
Predetermined outcomes for all students	Conceptually equivalent outcomes, variation in "surface" forms
Organization	
Hierachical structure, principal as manager	Mutual decisions, principal as head teacher
Individual work by isolated teachers	Professional community of inquiry
Separate grade levels; pull-out programs and specialists to handle problem cases	Upgraded adaptations, school-wide integrated services

Figure 2. Contrast between Factory and Information-Age models of schooling.

political concern; to remain economically competitive, U. S. schools need to provide for virtually all students a level of education previously limited to the privileged elite. This goal is all the more daunting given dramatic increases during the past two decades in the proportion of children living in poverty.

The second development is the evolution of the cognitive model described earlier. This model, which has seen application to curriculum and instruction only in limited "laboratory" settings, emphasizes reflection, strategic process-oriented learning, and social constructivism, all of which are foreign ideas and practices for most teachers, all of which are difficult to "package." Understanding the implications of the information-age model for schools therefore requires close attention (a) to curriculum and instruction, (b) to the teacher's role, and (c) to assessment.

A cognitive approach to *curriculum*, the development of a "curriculum of thoughtfulness," builds on assumptions quite different from the behavioral model:

- The mind is a living organ that depends on purpose and coherence, not a warehouse to be filled with information.
- Reflective learning built on genuine dialogue and social interaction is more long-lasting and transferable than rote acquisition.
- Previous experience is essential for effective learning.

Several cognitive psychologists, including my colleagues and me, have developed curriculum programs that incorporate these principles (Calfee, in press). Our work has focused on professional development; others favor packaged materials or computer software. We have been guided by "three C's"—coherence, connectedness, and communication (Figure 3). *Coherence* refers to the limits of short-term attentional memory, which we concretize in an aphorism: "KISS The Turkey!" The K.I.S.S. principle comes from Peters and Waterman, *In Search of Excellence* (1982), who found that successful businesses "Keep it simple, sweetheart!" But how does this principle apply to the classroom teacher, for whom the basal reader is the ultimate in intricacy with its thousands of objectives? How to simplify complexity?

The answer is that "Simple isn't easy." We liken the K.I.S.S. task to carving a turkey; unless you have x-ray vision to see the joints, you are likely to make hash. Whether for the entire curriculum from kindergarten through sixth grade, for a thirty-minute lesson, or for a three-week project, the key is to divide the whole into a small number of chunks. Otherwise the result is a "lump" (an indigestible blob) or a "mess" (a chaotic collection of factoids). Using such metaphors to translate from theory to practice may appear simpleminded, but it works. In place of basalized lessons with a multitude of tidbits, teachers are freed to design lessons around a few interrelated concepts.

Connectedness refers to the linkage between prior experience and new learning. Given the incredible diversity of today's students, this task appears

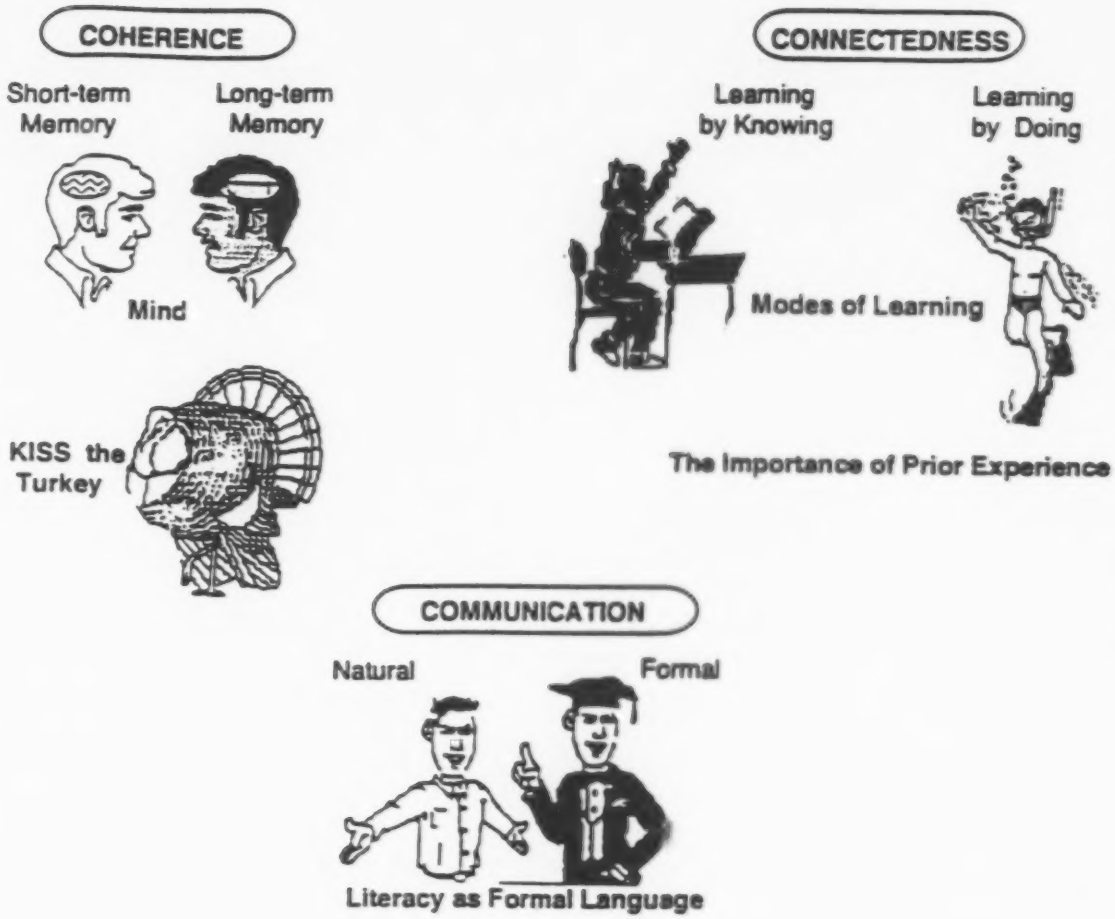

Figure 3. Conceptual elements for a cognitive model of literacy.

at first an incredible challenge. It is understandable that teachers sometimes throw up their hands in despair; "These kids don't know anything—they watch too much television." The key is to link to what students *do* know, rather than emphasizing what they *do not* know. Children from disadvantaged backgrounds may not be familiar with "school book" knowledge, but they have a wealth of information about the world, much of it outside the scope of the teacher's experience.

Communication in this model emphasizes the distinction between natural and formal language (Goody, 1987). This contrast, springing from psycholinguistics and cultural anthropology, assumes that all children enter school with a fully functioning linguistic system, but that they vary in the natural language acquired during childhood and in their familiarity with the formal language that is the standard for school and for society.

Formal language contrasts in several ways with natural language. First, in this definition literacy has less to do with *medium* and *message* than *manner*. When elementary teachers talk about "learning to read," they usually mean that the student can real aloud, can decode the printed text. They equate reading with textbooks; the fifth grader who has a paperback in his pocket (and who may commit occasional graffiti) is illiterate if he neglects the assigned social studies chapter on American Indians.

Communication also includes *meta-talk* as an essential ingredient for critical literacy. Cognitive psychologists use *meta-cognition* to describe "talking about thinking," a concept that is inherently social and communicative. The human capacity to reflect is uniquely linked to language, but is not an automatic consequence of linguistic competence. Vygotsky (1962) argued persuasively that reflectiveness emerges through a developmental progression beginning with the egocentric preschooler's efforts to be understood by others, leading eventually to the capacity to understand himself or herself.

The explicitness of formal language thereby links to the *social dimension* of critical literacy. In everyday usage, *criticism* implies harsh judgments; for the Greeks, however, a critic was an individual who could explain and judge the merits and shortcomings of an event or object, a connoisseur. *Functional literacy* allows a person to use language to do something—to read a want ad or use a technical manual to fix a leaky sink. *Critical literacy* includes the capacity for action, but incorporates a broader sense of understanding and insight, and the ability to communicate with others about "texts," both written or spoken. It is the difference between understanding how to operate the lever in a voting booth versus deciding for whom to vote and why.

In short, we conceive a cognitive curriculum that relies on a "deep" rather than "surface structure" definition of "what should be learned." The *critical literacy* model emphasizes acquiring strategic rather than content knowledge, collaborative rather than competitive learning. It includes elements of Socratic dialogues, a dash of meta-cognitive strategies, strong reliance on the wisdom of practice, and reliance on available resources within the classroom rather than pleas for "more materials."

The information-age model also entails a dramatic shift in the *teacher's role*, in the way that teachers think and act as individuals and as collaboratives. Michael Apple (1990) and others have blamed the spread of the factory model throughout U. S. schools for the "de-skilling" or deprofessionalization of teachers. This shift shows up when teachers ask, "Why don't you just tell us what you want us to do?" or "I'm not sure that 'they' will let us do that." Changing teacher cognitions is a substantial challenge if U. S. schools are to achieve authentic "cognitive" education. The information-age paradigm entails significant changes in institutional arrangements for teachers and principals; it is inconceivable that classrooms can operate in an information age while the school continues the factory tradition. As Sarason (1990) put it, "Whatever factors, variables, and ambience are conducive to the growth, development, and self-regard of a school's staff are precisely those that are crucial for obtaining the same consequences for students in a classroom" (p. 152).

Figure 2 also points the direction for reform in *assessment*. Some of the proposed changes are "ahead to the past," in that they bear a striking resemblance to examinations employed by teachers in the years before standardized tests. The United States is alive with the ferment of "alternative assessment"; piles of articles scattered around my study herald the latest ideas about *authentic assessment, performance and projects, exhibitions, portfolios,* and so on. Many of these activities have their origin in teachers' dissatisfaction with standardized methods, with their search for legitimization of their capacity to judge student achievement (Hiebert & Calfee, 1992). The same teachers often hearken to new trends in curriculum and instruction—whole language, process writing, cooperative learning.

Here are some data that inform current developments in alternative assessment. Under auspices of the *National Center for the Study of Writing,* Pam Perfumo and I have conducted a nation-wide survey of portfolio practice (Calfee & Perfumo, 1993). Our goal was to move beyond headlines (and newsletter reports) toward a deeper portrayal of what educators mean when they say that they are "doing portfolios." The survey focused on *writing* assessment, but the products were equally often linked to reading instruction.

The survey, which included 150 "nominated" contacts, including states, districts, schools, school teams, and individual teachers, was not random, but rather aimed to assess best practice. To guide the respondents (and to structure the responses), we divided the survey into "chunks": *Background and History* (how did you get into portfolios?); *Portfolios in the Classroom* (what does the concept mean in practice?); *Portfolio Process* (how do you do it?); *Portfolio Impact* (what do you see as the effect of portfolios for your students and for you?).

Our analyses turned up three themes: (a) teachers enlisted in the portfolio movement convey an intense commitment and personal renewal; (b) the technical foundations for portfolio assessment appear infirm and inconsistent at all levels; and (c) portfolio practice at the school and teacher level shies away from standards and grades, toward narrative and descriptive reporting.

First, commitment and renewal. Across wide variations in approaches and definition, the portfolio approach has energized the professional status and development of educators, especially classroom teachers. This response is partly affective; people who previously viewed themselves as a subclass tell about spending enormous amounts of time and energy rethinking the meaning of their work, and they feel invigorated by a renewed commitment. A common theme is "ownership." Teachers talk about "being in charge" of instruction. They describe the benefit to students who take responsibility for assessing their own writing.

Second, the surveys, interviews, and documents all disclose a lack of analytic and technical substance. For instance, respondents claim that an important purpose of portfolios is valid assessment of student progress and growth, yet nowhere in the packets have we found a clear account of how achievement is to be measured. District and state activities generally attempt to incorporate judgments and standards, usually through holistic ratings by external evaluators; school and classroom projects less often describe how to convert a folder of work into an achievement indicator. The procedures are normative rather than developmental. Also missing is discussion of conventional (or unconventional) approaches for establishing validity and reliability. Validity is assumed to inhere in the authenticity of the portfolio process; reliability is simply not discussed.

Third, respondents exhibited a definite distaste for evaluation. They do not want to set standards or assign grades for students or programs. This reaction is captured by the remark, "I wish grades would just go away!" Teachers are willing to judge individual compositions and other student work samples, but uncomfortable about assessing an entire portfolio.

Nowhere in the array of data did we find evidence for the impact of principles from either cognitive psychology or psychometrics! Teachers and administrators are guided by the pragmatics of schooling and the intuitions of their craft. The current reform is not so much a paradigm shift as a "workers' revolt." The teachers' goal is partly to alleviate the stultifying boredom of text-book-driven instruction, but their basic thrust is "Leave me alone with my kids and I'll do the best I can—trust me!"

External and Internal Mandates

The contrast between bottom-up activities described at the end of the previous section and the top-down efforts of policy-makers leads me to the following questions for this third episode:

- Who is in charge of assessment?
- Who is going to be affected by the results?
- What are the stakes?

The struggle to find answers to these questions cuts to the core of educational policy and practice in the United States and, I suspect, many other places.

They are important for students ("Is this going to be on the test?"). They are important for teachers, as shown by surveys of the impact of high-stakes tests on curriculum, instruction, attitudes, and ploys ("What do we do to raise test scores?). And they are important for policy makers; WYGIWYT, "What you get is what you test," is presently driving the U. S. toward a large system of voluntary national tests and associated standards (Shepard, 1992).

Important though they may be, these questions do not directly connect with issues of educational reform. How to "do it" and whether "it" is behavioral or cognitive means little in policy discussions. The basic tensions are portrayed in Figure 4. Bridging this gap is perhaps the most significant task confronting U. S. educators. As long as externally-mandated instruments are "what counts," the cognitive revolution is unlikely to have much impact on most classrooms. Moreover, the schools most impacted by the factory model are those serving children often at risk for school failure because of family circumstances; they are most likely to be "managed."

The external approach has a well-defined technology in psychometrics in the standardized test model. Nature abhors a vacuum, and so, unsurprisingly, standardized assessment techniques have appeared in classrooms under the rubric of *measurement-driven instruction*. Hiebert and I (Calfee & Hiebert, 1991; also see Cronbach, 1960) have proposed an alternative model of classroom-based assessment as a form of applied social science research. The teacher-based research perspective takes shape as a set of practical questions:

- *Purpose* (What are the goals? What working hypotheses guide the activity?)
- *Method* (How should the data be collected? How should the inquiry be designed?)
- *Interpretation and reporting* (Is the evidence reliable? Valid? What does it mean? What are the options for action?).

Implicit in this model is the ideal of a thoughtful, cognitively-oriented teacher. But can "regular" classroom teachers really be trusted with the challenge of defining high-level achievement outcomes, identifying or constructing authentic assessment tasks for these outcomes, and evaluating those tasks? The conceptual base is complex, requiring knowledge of the reading and writing curriculum and instruction, as well as assessment strategies. Most U. S. teachers received their pre-service training a decade ago or more, and the evidence suggests that this preparation was often brief and unrelated to classroom assessment or instructional practice (Stiggins, Conklin, & Bridgeford, 1986). Surveys of teacher-based assessment turn up haphazard collections of student work and poorly constructed performance-based assessments. Teachers appear ill-equipped and feel unable to handle the challenge of authentic assessment. Although I think that teachers actually have the potential to meet the challenge, they will need well-designed and adequately supported staff development in classroom assessment. Moreover,

Comparison between Assessment Instruments
Designed for Different Purposes

Assessment Designed for Instruction	Assessment Designed for External Accountability
Purpose and Source	
Teacher designed for classroom decisions	Designed by experts for policy makers
Combines several sources of information	Stand-alone, single index
Strong link to curriculum and instruction	Independent of curriculum and Instruction
Criteria	
Valid for guiding instruction	Predictive validity
Profile reliability--strengths and weaknesses	Total test reliability
Sensitive to dynamic changes in performance	Stable over time and situations
Performance is often all-or-none	Normally distributed scores
Pragmatics	
Judgmental, quick turn-around, flexible	Objective, cost and time efficient, standardized
Performance-based, "real" task	Multiple-choice, recognition
Administer whenever needed	Once-a-year, sometimes twice

Figure 4. Contrasts between internally-and externally-mandated concepts of testing and assessment.

such staff development must connect with the pragmatics of validity and reliability. Authentic assessment promises validity, but technical support for this claim is another matter. Face validity (does the "test" resemble what it claims to assess?) is assumed in authentic assessment, but it is often "activity-based" rather than conceptually grounded. Construct validity is the greatest challenge for any assessment; the potential of alternative methods, including portfolios, depends on strengthening the linkages to curriculum and instruction, and developing effective techniques for analysis, interpretation, and reporting.

Reliability is another matter. Although proponents of alternative assessment stake many of their claims on the validity of the tasks, few address reliability: consistent interpretation of student work over judges and tasks, and generalizability across contexts. Variability in tasks and contexts is expected in authentic assessment, further complicating the reliability issue.

Next is the issue of standards and criteria for judging the quality of student work. Researchers are confronted with the task of interpreting findings and making decisions about the significance of an outcome; so must the teacher as researcher. Collecting and reviewing work samples is engaging, even compelling; evaluating strengths and weaknesses is more difficult, but essential for assessment.

Finally, assessment results must be communicated to others. Authentic assessment is demanding; it requires expertise, time, and commitment. Many U. S. teachers endorse the concept because it is consistent with contemporary views of reading and writing, but most will not sustain the extra burdens unless others, outside the classroom, understand and value the information. The challenge is to communicate with a diverse audience of parents, administrators, concerned citizens, and policy makers, while maintaining the integrity and instructional value of authentic assessment.

How Will This Story End?

Newton showed that inertia is a powerful principle in the physical world, and the same seems to hold in the psychological and social arenas. Predicting the state of U. S. schools a decade from now, the best guess would be, "Pretty much as they are now." Which is actually not as bad as some people say, all things considered, but schools do need to improve.

Another scenario, favored by some cognitive scientists, replaces teachers with technology. This strategy seems unlikely if we are talking about "children and youth," youngsters between five and fifteen, kindergartners and adolescents. While modern technology can support teachers' efforts, effective education of students within this age range needs to be people-oriented more than machine-oriented. Good teachers are especially critical for students who lack social models and support for schooling at home.

What does contemporary cognitive psychology have to say about assessment and instruction under these conditions? The field has several points to make. For example, both assessment and instruction must be

contextualized, reflective, social. A major thesis of the new generation of Cognitivists is the importance of ecological *validity*. Laboratory findings have been criticized for their artificiality, and the same holds for applied cognition. It is easy to find situations in which students fail; what we need to create are "clean tests" eliminating unnecessary barriers to success.

A strategy for achieving this goal relies on the teacher to serve as a trustworthy judge for gauging student achievement, taking into account the setting for instruction, the setting for assessment, and the need to "experiment." A cognitive curriculum requires a thoughtful teacher, and a valid assessment demands professional judgments. Within this framework, portfolio collections of student work serve a function, but they need to be analyzed and interpreted by the teacher. How to deal with issues of reliability and trustworthiness? How to connect with other assessment methods and outcomes (e.g., grades, parent conferences, standardized tests)? How to manage consistency for students during their years of schooling within and between grades and schools?

In the U. S., the most serious hurdle in the way of implementing the preceding concepts and answering the previous questions is the difficulty of sustaining systematic teacher assessment. On the surface, collecting student work is simple enough; difficulties arise in deciding how to select work samples and how to assess these samples in an informative and consistent manner. My colleagues and I have developed the concept of the *Teacher Logbook* to address these issues. Figure 5 shows how the Logbook can accomplish three interrelated tasks: *documentation* of evidence bearing on student performances; *summary judgments* of student achievement; and a *curriculum record*.

Critical to the Logbook technique is the concept of a developmental curriculum, a small set of critical domains with mileposts that serve as targets for the school. For instance, in the literacy curriculum, comprehension and composition in the narrative genre is an important outcome for the elementary grades. Within the narrative form, for example, four outcomes are generally recognized as critical for competence in handling literature: character, plot, setting, and theme. For kindergartners, appreciating the moral of simple fables may be a reasonable goal. By second grade, students may be expected to identify thematic issues implicit in a work such as *Charlotte's Web*, and to express the meaning of the work in personal terms. Sixth graders should be fully capable of employing thematic elements in their own compositions, and to identify multiple themes in collections of related texts.

As laid out in the figure, student summaries are placed at the beginning of the Logbook, because these play the most critical role in reporting student achievement. We imagine a procedure in which, on a regular basis, perhaps once a quarter, the teacher conducts a formal rating of each student's achievement level in the Summary section of the Logbook. The entries reflect the teacher's judgment about each student's location on the developmental curriculum scale. For instance, a teacher might judge a third grade student as handling theme like a first grader, still at the level of mundane morals.

THE TEACHER LOGBOOK

Section I: Student Summary

Fall Entry Level

Student	Reading/Writing/Language				Math . . .
	Vocab	Narrative	Expos	Skills	
Able, J.					
. . .					
Zeno, K.					

Section II: Journal Notes

Week of _____

Section III: Curriculum Plan/Record

Plans for Fall Qtr

Sept:	Activities	Vocab Narr Expos Skills
	Update	
Dec:	Activities	Vocab Narr Expos Skills
	Update	

Figure 5. Design of a Teacher Logbook for documenting and summarizing the teacher's assessment of student achievement.

The journal in the middle of the Logbook provides space for the teacher to record ongoing information relevant to student performance: observations, informal assessments of student activities and projects, and questions requiring further thought and action. The notes are a natural place for comments about student portfolio entries, along with more formal assessments. Curriculum planning is at the end of the Logbook. These entries are quite different from the routinized "lesson plans" typically completed by teachers to meet bureaucratic mandates. They are long-term working plans organized by curriculum goals, with room for commentary and revision.

The Logbook concept builds on the notion that the teacher, with a developmental curriculum in mind, regularly records brief notes about individual students in the "profile" section. The comments provide a concrete record for reflection and action. An empty profile sheet is a reminder that the student has slipped from sight. A sheet showing a long list of "books read" but no evidence of written work is a prod to encourage the student to put his or her thoughts on paper. Teachers keep mental records of this sort; the Logbook is designed as a "memory jogger," and a source of information for reflection and assessment.

The Logbook also provides a methodology for addressing issues of validity and reliability: How can the teacher's summary judgments about students be gauged for consistency and trustworthiness? My answer to this question relies on the concept of *panel judgments;* much like an Olympic panel, classroom teachers can validate their evaluations through cross-checks (the British refer to this process as the "moderation" task). Again, the workability of this approach relies on the emergence of the teacher as a practical researcher within a school that provides a context for assessment. Several examples can be found to support the practicality of this proposal. In California, for example, panels are incorporated in the Self-Study and PQR (Program Quality Review) process conducted by every school in the state once every three years. The idea is also reflected in the frameworks produced by professional organizations (e.g., NCTE and IRA), in the work of grade level teams in many elementary schools, in the maintenance of department standards in secondary schools, and in the shared leadership typical of school restructuring.

Conceptually, the panel-judgment process can call upon established methods of generalizability theory as a foundation (Shavelson & Webb, 1991). To be sure, application of the theory to panel judgments requires the construction of designs that identify significant factors likely to influence the judgment process. As a first cut, we suggest as critical factors the *curriculum domain* (holistic assessment of an entire portfolio is likely to fall prey to the same variability as for writing samples; the survey teachers were wise when they resisted holistic judgments), *task conditions* (e.g., standardized vs. open-ended, constrained vs. project-based), *contextual factors* (e.g., individual vs. group, with or without instructional support and resources), and *characteristics of the judges* (e.g., colleagues, administrators, external experts).

The conceptual task of designing and validating the Logbook concept is no less demanding than the practical issues of implementation. The survey responses show little evidence of systematic documentation by teachers, unless this action was externally mandated. Wolf's (1992) dissertation on classroom portfolios (similar to the Logbook) is rich in its accounts of student work samples, but thin on teacher records. Teachers agreed to document the performance of two target students, but ran out steam midway through the school year. In Shulman's (1990) Teacher Assessment Project, teacher logs were an important component in the design of the Literacy component. Beginning teachers compiled professional portfolios during the school year for display during a performance demonstration before an expert panel comprising peers and academics. Collegial meetings during the year provided direction and support. The candidates, third grade teachers, included in their professional portfolio a progress record (akin to the Logbook) for four target students within their classroom. The results showed that, given adequate support and purpose, teachers found the documentation task both feasible and informative. Let me suggest that the teacher logbook also offers a technique for preparing teachers in assessment technology—not in classical psychometrics, but in the conceptual pragmatics of psychometric principles: convergent validity and faceted consistency.

Alternative assessment and student portfolios tend to appear in combination with other elements: whole language rather than basal readers, cooperative instruction rather than didactic teacher-talk, school-based decision-making rather than top-down direction, the teacher as professional rather than as civil servant. My sense is that such strategies offer the opportunity for fundamental reform in U. S. schooling. Reform efforts are presently piecemeal and unrelated, overwhelming teachers by a multiplicity of demands. The enthusiasm and commitment of portfolio teachers are impressive, but the high costs and limited benefits are discouraging. The portfolio movement seems likely to falter and fail unless it is connected to the other supporting components in a manner that continues to meet internal classroom needs (valid data for instructional decisions) while satisfying external policy demands (reliable information for accountability purposes). The Teacher's Logbook is a bridge for spanning this chasm. For the Logbook to become a reality will require (a) establishment of a serious "audience" for this activity, and (b) provision of adequate professional development. And if alternative assessment methods are to realize their full potential, they must be connected to curriculum and instruction that embodies the cognitive principles appropriate to an information-age schooling. Notice that I am not calling for the abolition of externally-mandated tests, but for the elevation of information from internally-mandated assessments to a complementary level—to equal status for significant policy audiences.

Absent such support, my guess is that the portfolio movement will eventually fall of its own weight. Selected teachers will continue to rely on their professional judgment for deciding what to teach and how to teach it,

and for rendering assessments to interested audiences. External authorities may entertain the idea of portfolios, performances, and exhibitions, but cost-effectiveness will eventually carry the day (this shift has happened in the past; witness the early years of NAEP). And another chance to improve the quality of schooling in the United States will have slipped through our fingers. But I am an optimist. The convergence over the past 50 years of cognitive theory and research, more far-reaching psychometrics, and a renewed understanding of practical professionalism—this convergence leaves me hopeful!

REFERENCES

Apple, M. (1990). *Ideology and curriculum* (2nd ed.). New York: Routledge.

Berliner, D. A., & Calfee, R. C. (in press). *Handbook of educational psychology*.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32-42.

Brownell, W. A. (1948). Criteria of learning in educational research. *Journal of Educational Psychology, 39*, 170-182.

Calfee, R. C. (1981). Cognitive psychology and educational practice. In D. C. Berliner (Ed.), *Review of research in education*. Washington, D. C.: American Educational Research Association.

Calfee, R. C. (in press). Critical literacy: Reading and writing for a new generation. In C. Hedley & N. Ellsworth (Eds.), *Education for the future: Literacy perspectives*. Hillsdale NJ: Erlbaum.

Calfee, R. C., & Hiebert, E. H. (1991). Classroom assessment of reading. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of research on reading* (2nd ed., pp. 281-309). New York: Longman.

Calfee, R. C., & Perfumo, P. (1993). Student portfolios: Opportunities for a revolution in assessment. *Journal of Reading, 36*, 532-537.

Chi, M. T., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. J. (1960). *Essentials of psychological testing*. New York: Harper & Row.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*, 116-127.

Estes, W. M. (1980). Is human memory obsolete? *American Scientist, 68*, 62-69.

Gage, N. L. (Ed.). (1963). *Handbook of research on teaching*. Chicago: Rand-McNally.

Goody, J. (1987). *The interface between the written and the oral*. Cambridge, England: Cambridge University Press.

Greeno, J. G. (1980). Psychology of learning, 1960-1980: One participant's observations. *American Psychologist, 35*, 713-728.

Hiebert, F., & Calfee, R. C. (1992). Assessment of literacy: From standardized tests to performances and portfolios. In A. E. Farstrup and S. J. Samuels (Eds.), *What research says about reading instruction* (pp. 70-100). Newark, DE: International Reading Association.

Mecklenburger, J. A. (1992). The braking of the "Break the mold" express. *Phi Delta Kappan, 74*, 280-289.

Novick, L. R. (1990). Representational transfer in problem solving. *Psychological Science, 1*, 128-132.

Perkins, D. N., & Salomon, G. (1988). Teaching for transfer. *Educational Leadership, 45* (9), 22-32.

Peters, T. J., & Waterman, R. H., Jr. (1982). *In search of excellence*. New York: Harper and Row.

Sarason, S. B. (1990). *The predictable failure of school reform*. San Francisco: Jossey-Bass.

BEST COPY AVAILABLE

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park: Sage.

Shepard, L. A. (1992). Psychometricians' beliefs about learning. *Educational Researcher, 20*, 2-16.

Shulman, L. S. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership, 46*, 36-41.

Snow, R. E., & Jackson, D. N. (1992). Individual differences in conation: Selected constructs and measures.

Snow, R. E. (in press). Construct validity and constructed response tests.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement, 5*, 5-17.

Travers, R. M. (Ed.). (1973). *Second handbook of research on teaching*. Chicago: Rand-McNally.

Vygotsky, L. Y. (1962). *Thought and language*. Cambridge, MA: The MIT Press. Also by New York: Wiley.

Wittrock, M. C. (Ed.). (1986). *Handbook of research on teaching* (3rd ed.) New York: Macmillan.

Wolf, K. P. (1992). *Informed assessment of students through the classroom literacy portfolio*. Unpublished doctoral dissertation, Stanford University.

NATIONAL CENTER FOR THE STUDY OF WRITING
5513 Tolman Hall • Graduate School of Education
University of California • Berkeley, CA 94720
(510) 643-7022

The National Center for the Study of Writing, one of the national educational research centers sponsored by the U.S. Department of Education's Office of Educational Research and Improvement, is located at the Graduate School of Education at the University of California at Berkeley, with a site at Carnegie Mellon University in Pittsburgh, Pennsylvania. The Center provides leadership to elementary and secondary schools, colleges, and universities as they work to improve the teaching and learning of writing. The Center supports an extensive program of educational research and development in which some of the country's top language and literacy experts work to discover how the teaching and learning of writing can be improved, from the early years of schooling through adulthood. The Center's four major objectives are: (1) to create useful theories for the teaching and learning of writing; (2) to understand more fully the connections between writing and learning; (3) to provide a national focal point for writing research; and (4) to disseminate its results to American educators, policy-makers, and the public. Through its ongoing relationship with the National Writing Project, a network of expert teachers coordinated through Berkeley's Graduate School of Education, the Center involves classroom teachers in helping to shape the Center's research agenda and in making use of findings from the research. Underlying the Center's research effort is the belief that research both must move into the classroom and come from it; thus, the Center supports "practice-sensitive research" for "research-sensitive practice."

Sarah Warshauer Freedman, *University of California at Berkeley, **Director***
Anne Haas Dyson, *University of California at Berkeley, **Co-Director***
Linda Flower, *Carnegie Mellon University, **Co-Director***
James Gray, *University of California at Berkeley, **Co-Director***
J. R. Hayes, *Carnegie Mellon University, **Co-Director***
Donald McQuade, *University of California at Berkeley, **Professional and Community Liaison***
Sandra R. Schecter, *University of California at Berkeley, **Associate Director***

NATIONAL ADVISORY BOARD
Fred Hechinger, *Senior Advisor, Carnegie Corporation of New York, **Co-Chair***
Courtney Cazden, *Professor, Harvard University, **Co-Chair***

Marcia Farr, *Professor, University of Illinois, Chicago*
Phyllis Franklin, *Executive Director, Modern Language Association*
Erminda Garcia, *Teacher, Hall District Elementary School, Watsonville, California*
Sibyl Jacobson, *Executive Director, Metropolitan Life Foundation*
Alice Kawazoe, *Director of Staff and Curriculum Development, Oakland Unified School District*
Luis C. Moll, *Associate Professor, University of Arizona*
Miles Myers, *Executive Director, National Council of Teachers of English*
Yolanda Peeks, *Principal, Brookfield Elementary School, Oakland, California*
Stan Pesick, *Teacher, Skyline High School, Oakland, California*
Jerrie Cobb Scott, *Director, Center for Studies of Urban Literacy, Central State University, Wilberforce, Ohio*
Lee Shulman, *Professor, Stanford University*
Carol Tateishi, *Director, Bay Area Writing Project*